09/786009
10 Rec'd PCT/PTO 17 APR 2001
PCT09

```
RAW SEQUENCE LISTING                          DATE: 09/27/2001
PATENT APPLICATION:  US/09/786,009            TIME: 12:11:57

Input Set : A:\Inteinm.app
Output Set: N:\CRF3\09272001\I786009.raw

      3 <110> APPLICANT: XU, Ming-Qun
      4       EVANS, Thomas C.
      6 <120> TITLE OF INVENTION: Intein Mediated Peptide Ligation
      8 <130> FILE REFERENCE: NEB-150PUS
     10 <140> CURRENT APPLICATION NUMBER: 09/786,009
C--> 11 <141> CURRENT FILING DATE: 2001-04-17
     13 <150> PRIOR APPLICATION NUMBER: 60/102,413
     14 <151> PRIOR FILING DATE: 1999-09-30
     16 <150> PRIOR APPLICATION NUMBER: PCT/US99/22776
     17 <151> PRIOR FILING DATE: 1999-09-30
     19 <160> NUMBER OF SEQ ID NOS: 9
     21 <170> SOFTWARE: PatentIn Ver. 2.0
     23 <210> SEQ ID NO: 1
     24 <211> LENGTH: 43
     25 <212> TYPE: DNA
     26 <213> ORGANISM: Artificial Sequence
     28 <220> FEATURE:
     29 <223> OTHER INFORMATION: Description of Artificial Sequence: the modified
     30       C-terminal splice junction of the intein from the
     31       gyrA gene of Mycobacterium xenopi
     33 <400> SEQUENCE: 1
     34 ggttcgtcag ccacgctact ggcctcaccg gttgatagct gca                      43
     36 <210> SEQ ID NO: 2
     37 <211> LENGTH: 39
     38 <212> TYPE: DNA
     39 <213> ORGANISM: Artificial Sequence
     41 <220> FEATURE:
     42 <223> OTHER INFORMATION: Description of Artificial Sequence: the
     43       complementary strand of the C-terminal splice
     44       junction of the modified intein from the gyrA
     45       gene of Mycobacterium xenopi
     47 <400> SEQUENCE: 2
     48 gctatcaacc ggtgaggcca gtagcgtggc tgacgaacc                           39
     50 <210> SEQ ID NO: 3
     51 <211> LENGTH: 68
     52 <212> TYPE: DNA
     53 <213> ORGANISM: Artificial Sequence
     55 <220> FEATURE:
     56 <223> OTHER INFORMATION: Description of Artificial Sequence: the polylinker
     57       sequence inserted upstream of the modified intein
     58       from the gyrA gene of Mycobacterium xenopi
     60 <400> SEQUENCE: 3
     61 tttaatctag acatatggcc atgggtggca gccgcctcga gggctcttcc tgcatcacgg 60
     62 gagatgca                                                             68
     64 <210> SEQ ID NO: 4
     65 <211> LENGTH: 69
     66 <212> TYPE: DNA
```

ENTERED

```
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: At position 41, "H" = A or C or T.
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: the
      complementary strand of the polylinker inserted
      upstream of the modified intein from the gyrA
      gene of Mycobacterium xenopi
<400> SEQUENCE: 4
ctagtgcatc tcccgtgatg caggaagagc cctcgaggcg hgccgccacc catggccata 60
tgtctagat                                                         69
<210> SEQ ID NO: 5
<211> LENGTH: 6509
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: pTXB1 plasmid
      sequence containing the modified intein from the
      gyrA gene of Mycobacterium xenopi
<400> SEQUENCE: 5
aactacgtca ggtggcactt ttcggggaaa tgtgcgcgga acccctattt gtttattttt 60
ctaaatacat tcaaatatgt atccgctcat gagacaataa ccctgataaa tgcttcaata 120
atattgaaaa aggaagagta tgagtattca acatttccgt gtcgccctta ttcccttttt 180
tgcggcattt tgccttcctg tttttgctca cccagaaacg ctggtgaaag taaaagatgc 240
tgaagatcag ttgggtgcac gagtgggtta catcgaactg gatctcaaca gcggtaagat 300
ccttgagagt tttcgccccg aagaacgttt tccaatgatg agcactttta aagttctgct 360
atgtggcgcg gtattatccc gtgttgacgc cgggcaagag caactcggtc gccgcataca 420
ctattctcag aatgacttgg ttgagtactc accagtcaca gaaaagcatc ttacggatgg 480
catgacagta agagaattat gcagtgctgc cataaccatg agtgataaca ctgcggccaa 540
cttacttctg acaacgatcg gaggaccgaa ggagctaacc gcttttttgc acaacatggg 600
ggatcatgta actcgccttg atcgttggga accggagctg aatgaagcca taccaaacga 660
cgagcgtgac accacgatgc ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg 720
cgaactactt actctagctt cccggcaaca attaatagac tggatggagg cggataaagt 780
tgcaggacca cttctgcgct cggcccttcc ggctggctgg tttattgctg ataaatctgg 840
agccggtgag cgtgggtctc gcggtatcat tgcagcactg gggccagatg gtaagccctc 900
ccgtatcgta gttatctaca cgacggggag tcaggcaact atggatgaac gaaatagaca 960
gatcgctgag ataggtgcct cactgattaa gcattggtaa ctgtcagacc aagtttactc 1020
atatatactt tagattgatt taccccggtt gataatcaga aaagccccaa aaacaggaag 1080
attgtataag caaatattta aattgtaaac gttaatattt tgttaaaatt cgcgttaaat 1140
ttttgttaaa tcagctcatt ttttaaccaa taggccgaaa tcggcaaaat cccttataaa 1200
tcaaaagaat agcccgagat agggttgagt gttgttccag tttggaacaa gagtccacta 1260
ttaaagaacg tggactccaa cgtcaaaggg cgaaaaaccg tctatcaggg cgatggccca 1320
ctacgtgaac catcacccaa atcaagtttt ttggggtcga ggtgccgtaa agcactaaat 1380
cggaacccta aagggagccc ccgatttaga gcttgacggg gaaagccggc gaacgtggcg 1440
agaaaggaag ggaagaaagc gaaaggagcg ggcgctaggg cgctggcaag tgtagcggtc 1500
acgctgcgcg taaccaccac acccgccgcg cttaatgcgc cgctacaggg cgcgtaaaag 1560
gatctaggtg aagatccttt ttgataatct catgaccaaa atcccttaac gtgagttttc 1620
gttccactga gcgtcagacc ccgtagaaaa gatcaaagga tcttcttgag atcctttttt 1680
tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg ctaccagcgg tggtttgttt 1740
```

```
RAW SEQUENCE LISTING                          DATE: 09/27/2001
PATENT APPLICATION:  US/09/786,009            TIME: 12:11:57

Input Set : A:\Inteinm.app
Output Set: N:\CRF3\09272001\I786009.raw

122 gccggatcaa gagctaccaa ctctttttcc gaaggtaact ggcttcagca gagcgcagat 1800
123 accaaatact gtccttctag tgtagccgta gttaggccac cacttcaaga actctgtagc 1860
124 accgcctaca tacctcgctc tgctaatcct gttaccagtg gctgctgcca gtggcgataa 1920
125 gtcgtgtctt accgggttgg actcaagacg atagttaccg gataaggcgc agcggtcggg 1980
126 ctgaacgggg ggttcgtgca cacagcccag cttggagcga acgacctaca ccgaactgag 2040
127 atacctacag cgtgagctat gagaaagcgc cacgcttccc gaagggagaa aggcggacag 2100
128 gtatccggta agcggcaggg tcggaacagg agagcgcacg agggagcttc cagggggaaa 2160
129 cgcctggtat ctttatagtc ctgtcgggtt tcgccacctc tgacttgagc gtcgattttt 2220
130 gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc agcaacgcgg cctttttacg 2280
131 gttcctggcc ttttgctggc cttttgctca catgttcttt cctgcgttat cccctgattc 2340
132 tgtggataac cgtattaccg cctttgagtg agctgatacc gctcgccgca gccgaacgac 2400
133 cgagcgcagc gagtcagtga gcgaggaagc tatggtgcac tctcagtaca atctgctctg 2460
134 atgccgcata gttaagccag tatacactcc gctatcgcta cgtgactggg tcatggctgc 2520
135 gccccgacac ccgccaacac ccgctgacgc gccctgacgg gcttgtctgc tcccggcatc 2580
136 cgcttacaga caagctgtga ccgtctccgg gagctgcatg tgtcagaggt tttcaccgtc 2640
137 atcaccgaaa cgcgcgaggc agctgcggta aagctcatca gcgtggtcgt gcagcgattc 2700
138 acagatgtct gcctgttcat ccgcgtccag ctcgttgagt ttctccagaa gcgttaatgt 2760
139 ctggcttctg ataaagcggg ccatgttaag ggcggttttt tcctgtttgg tcacttgatg 2820
140 cctccgtgta agggggaatt tctgttcatg ggggtaatga taccgatgaa acgagagagg 2880
141 atgctcacga tacgggttac tgatgatgaa catgcccggt tactggaacg ttgtgagggt 2940
142 aaacaactgg cggtatggat gcggcgggac cagagaaaaa tcactcaggg tcaatgccag 3000
143 ccgaacgcca gcaagacgta gcccagcgcg tcggccgcca tgccggcgat aatggcctgc 3060
144 ttctcgccga aacgtttggt ggcgggacca gtgacgaagg cttgagcgag ggcgtgcaag 3120
145 attccgaata ccgcaagcga caggccgatc atcgtcgcgc tccagcgaaa gcggtcctcg 3180
146 ccgaaaatga cccagagcgc tgccggcacc tgtcctacga gttgcatgat aaagaagaca 3240
147 gtcataagtg cggcgacgat agtcatgccc cgcgcccacc ggaaggagct gactgggttg 3300
148 aaggctctca agggcatcgg tcgagatccc ggtgcctaat gagtgagcta acttacatta 3360
149 attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgtcgtgcca gctgcattaa 3420
150 tgaatcggcc aacgcgcggg gagaggcggt ttgcgtattg ggcgccaggg tggtttttct 3480
151 tttcaccagt gagacgggca acagctgatt gcccttcacc gcctggccct gagagagttg 3540
152 cagcaagcgg tccacgctgg tttgccccag caggcgaaaa tcctgtttga tggtggttaa 3600
153 cggcgggata taacatgagc tgtcttcggt atcgtcgtat cccactaccg agatatccgc 3660
154 accaacgcgc agcccggact cggtaatggc gcgcattgcg cccagcgcca tctgatcgtt 3720
155 ggcaaccagc atcgcagtgg gaacgatgcc ctcattcagc atttgcatgg tttgttgaaa 3780
156 accggacatg gcactccagt cgccttcccg ttccgctatc ggctgaattt gattgcgagt 3840
157 gagatattta tgccagccag ccagacgcag acgcgccgag acagaactta atgggcccgc 3900
158 taacagcgcg atttgctggt gacccaatgc gaccagatgc tccacgccca gtcgcgtacc 3960
159 gtcttcatgg gagaaaataa tactgttgat gggtgtctgg tcagagacat caagaaataa 4020
160 cgccggaaca ttagtgcagg cagcttccac agcaatggca tcctggtcat ccagcggata 4080
161 gttaatgatc agcccactga cgcgttgcgc gagaagattg tgcaccgccg ctttacaggc 4140
162 ttcgacgccg cttcgttcta ccatcgacac caccacgctg gcacccagtt gatcggcgcg 4200
163 agatttaatc gccgcgacaa tttgcgacgg cgcgtgcagg gccagactgg aggtggcaac 4260
164 gccaatcagc aacgactgtt tgcccgccag ttgttgtgcc acgcggttgg gaatgtaatt 4320
165 cagctccgcc atcgccgctt ccactttttc ccgcgttttc gcagaaacgt ggctggcctg 4380
166 gttcaccacg cgggaaaccg tctgataaga gacaccggca tactctgcga catcgtataa 4440
167 cgttactggt ttcacattca ccaccctgaa ttgactctct tccgggcgct atcatgccat 4500
168 accgcgaaag gttttgcgcc attcgatggt gtccggatct cgacgctctc ccttatgcga 4560
169 actcctgcat taggaagcag cccagtagta ggttgaggcc gttgagcacc gccgccgcaa 4620
170 ggaatggtgc atgccgccct ttcgtcttca agaattaatt cccaattcca ggcatcaaat 4680
```

```
aaaacgaaag gctcagtcga aagactgggc ctttcgtttt atctgttgtt tgtcggtgaa 4740
cgctctcctg agtaggacaa atccgccggg agcggatttg aacgttgcga agcaacggcc 4800
cggagggtgg cgggcaggac gcccgcccata aactgccagg aattaattcc aggcatcaaa 4860
taaaacgaaa ggctcagtcg aaagactggg cctttcgttt tatctgttgt ttgtcggtga 4920
acgctctcct gagtaggaca aatccgccgg gagcggattt gaacgttgcg aagcaacggc 4980
ccggagggtg gcgggcagga cgcccgccat aaactgccag gaattaattc caggcatcaa 5040
ataaaacgaa aggctcagtc gaaagactgg gcctttcgtt ttatctgttg tttgtcggtg 5100
aacgctctcc tgagtaggac aaatccgccg ggagcggatt tgaacgttgc gaagcaacgg 5160
cccggagggt ggcgggcagg acgcccgcca taaactgcca ggaattaatt ccaggcatca 5220
aataaaacga aaggctcagt cgaaagactg ggcctttcgt tttatctgtt gtttgtcggt 5280
gaacgctctc ctgagtagga caaatccgcc gggagcggat ttgaacgttg cgaagcaacg 5340
gcccggaggg tggcgggcag gacgcccgcc ataaactgcc aggaattggg gatcggaatt 5400
aattcccggt ttaaaccggg gatctcgatc ccgcgaaatt aatacgactc actatagggg 5460
aattgtgagc ggataacaat tcccctctag aaataatttt gtttaacttt aagaaggaga 5520
tatacatatg gctagctcgc gagtcgacgg cggccgcgaa ttcctcgagg gctcttcctg 5580
catcacggga gatgcactag ttgccctacc cgagggcgag tcggtacgca tcgccgacat 5640
cgtgccgggt gcgcggccca acagtgacaa cgccatcgac ctgaaagtcc ttgaccggca 5700
tggcaatccc gtgctcgccg accggctgtt ccactccggc gagcatccgg tgtacacggt 5760
gcgtacggtc gaaggtctgc gtgtgacggg caccgcgaac caccgttgt tgtgtttggt 5820
cgacgtcgcc ggggtgccga ccctgctgtg gaagctgatc gacgaaatca agccgggcga 5880
ttacgcggtg attcaacgca gcgcattcag cgtcgactgt gcaggttttg cccgcgggaa 5940
acccaatttt gcgcccacaa cctacacagt cggcgtccct ggactggtgc gtttcttgga 6000
agcacaccac cgagacccgg acgcccaagc tatcgccgac gagctgaccg acgggcggtt 6060
ctactacgcg aaagtcgcca gtgtcaccga cgccggcgtg caggcggtgt atagccttcg 6120
tgtcgacacg gcagaccacg cgtttatcac gaacgggttc gtcagccacg ctactggcct 6180
caccggtctg aactcaggcc tcacgacaaa tcctggtgta tccgcttggc aggtcaacac 6240
agcttatact gcgggacaat tggtcacata taacggcaag acgtataaat gtttgcagcc 6300
ccacacctcc ttggcaggat gggaaccatc caacgttcct gccttgtggc agcttcaatg 6360
actgcaggaa ggggatccgg ctgctaacaa agcccgaaag gaagctgagt tggctgctgc 6420
caccgctgag caataactag cataacccct tggggcctct aaacgggtct tgaggggttt 6480
tttgctgaaa ggaggaacta tatccggat                                   6509
<210> SEQ ID NO: 6
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: synthetic
      peptide
<400> SEQUENCE: 6
Cys Ala Tyr Lys Thr Thr Gln Ala Asn Lys His Ile Ile Val Ala Cys
 1               5                  10                  15
Glu Gly Asn Pro Tyr Val Pro Val His Phe Asp Ala Ser Val
            20                  25                  30
<210> SEQ ID NO: 7
<211> LENGTH: [illegible]
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: the amino acid
```

```
        sequence deduced from the polylinker region of
        pTXB1
<400> SEQUENCE: 7
Met Ala Met Gly Gly Gly Arg Leu Glu Gly Ser Ser Cys
  1               5                   10
<210> SEQ ID NO: 8
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: polylinker
        region upstream of the modified intein from the
        gyrA gene of Mycobacterium xenopi in pTXB1
<400> SEQUENCE: 8
catatggcca tgggtggcgg ccgcctcgag ggctcttcct gc                          42
<210> SEQ ID NO: 9
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: synthetic
        peptide
<400> SEQUENCE: 9
Cys Asp Pro Glu Lys Asp Ser
  1               5
```

```
VERIFICATION SUMMARY                              DATE: 09/27/2001
PATENT APPLICATION:  US/09/786,009                TIME: 12:11:58

Input Set : A:\Inteinm.app
Output Set: N:\CRF3\09272001\I786009.raw

L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date
```